THE SCIENTIST AS PROBLEM SOLVER

Technical Report AIP - 103

Herbert A. Simon

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1989

The Artificial Intelligence and Psychology Project

Departments of
Computer Science and Psychology
Carnegie Mellon University

Learning Research and Development Center
University of Pittsburgh

THE SCIENTIST AS PROBLEM SOLVER

Technical Report AIP - 103

Herbert A. Simon

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

1989

This research was supported by the Computer Sciences Division, Office of Naval Research, under contract number N00014-86-K-0678. Reproduction in whole or part is permitted for any purpose of the United States Government. Approved for public release; distribution unlimited.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AIP - 103	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Carnegie-Mellon University 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Computer Sciences Division Office of Naval Research
6c ADDRESS (City, State, and ZIP Code) Department of Psychology Pittsburgh, Pennsylvania 15213	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION Same as Monitoring Organization 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0678
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS p4000ub201/7-4-86 PROGRAM ELEMENT NO: N/A; PROJECT NO: N/A; TASK NO.: N/A; WORK UNIT ACCESSION NO: N/A

11 TITLE (Include Security Classification)

The Scientist as Problem Solver

12 PERSONAL AUTHOR(S)

Herbert A. Simon

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM 86Sept15 TO 91Sept14	1989	36

16 SUPPLEMENTARY NOTATION

Chapter 14 in D. Klahr and K. Kotovsky (Eds), COMPLEX INFORMATION PROCESSING

17 COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

problem solving
problem representation
scientific discovery
creativity
learning
induction of laws

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

SEE REVERSE SIDE

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Alan L. Meyrowitz	22b TELEPHONE (Include Area Code) (202) 696-4302 22c OFFICE SYMBOL N00014

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

ABSTRACT

Our exploration of the histories of scientific discoveries have made it eminently clear to us that scientists set themselves many different kinds of tasks. These include tasks of formulating significant scientific problems, of discovering interesting phenomena, of finding laws that are hidden in data (with and without the help of theories for guiding the search), or inventing new representations for phenomena and their accompanying theories, of inferring the logical consequences of theories and testing them.

The Scientist as Problem Solver

Herbert A. Simon
Carnegie-Mellon University

The thesis of this paper can be stated succinctly simply by replacing the "as" in its title by "is a" **The Scientist is a problem solver.** If the thesis is true then we can dispense with a theory of scientific discovery -- the processes of discovery are just applications of the processes of problem solving. However, since the thesis is not obvious to everyone, and since the topic of scientific discovery has interest in its own right, perhaps it is worth while saying a little more about it.

In a recent book (Langley, Simon, Bradshaw, and Zytkow, 1987), my co-authors and I have said a great deal more about discovery, and Deepak Kulkarni and I have added yet another chapter in a paper that has been submitted for journal publication (Kulkarni and Simon, 1986). There is no need to repeat these accounts here beyond the briefest summary of what we concluded and how we supported our conclusions. We concluded that the thesis is, indeed, valid. As evidence, we adduced careful reports of a substantial number of historical scientific discoveries, together with computer simulations that, starting with essentially the same initial conditions as did the human discoverers, made the same discoveries. Thus, the computer programs contained a set of processes that were sufficient for making the discoveries, and thereby provided a possible explanation for the success of the human scientists.

Our explorations of the histories of scientific discoveries have made it eminently clear to us that scientists set themselves many different kinds of tasks. These include tasks of formulating significant scientific problems, of discovering interesting phenomena, of finding the laws that are hidden in data (with and without the help of theories for guiding the search), of inventing new representations for phenomena and their accompanying theories, of inferring the logical consequences of theories and

testing them, of designing experiments, of finding explanatory mechanisms to account for empirical generalizations, and of inventing new instruments for observation and measurement. Undoubtedly there are others.

What is common to all of these tasks is that they appear to employ the same general kinds of problem solving processes as are employed by chessplayers in choosing moves, by subjects in the laboratory confronted with the Tower of Hanoi or the Missionaries and Cannibals problem, by physicians making diagnoses, by computer salesmen configuring systems for clients, by architects designing houses, or by organic chemists synthesizing new molecules. Mostly, they engage in heuristic search in a number of problem spaces: the spaces of theories and experiments mentioned by Klahr and Dunbar (this volume), but also spaces of problems, of phenomena, of representations, of instruments, and others.

Moreover, the "insight" that is supposed to be required for such work as discovery turns out to be synonymous with the familiar process of recognition; and other terms commonly used in the discussion of creative work -- such terms as "judgment," "creativity," or even "genius" -- appear either to be wholly dispensable or to be definable, as insight is, in terms of mundane and well understood concepts.

Until rather recent times, much of the published work about scientific discovery has consisted of anecdotes, frequently autobiographical, about specific discoveries and their finders. If discovery requires creativity, or even genius, it would be immodest for anyone to claim that he or she had made a discovery, and futile to try to describe how it had been done. But if discovery is plain, garden-variety problem solving, then there is no immodesty, and perhaps not even futility, in adding to the anecdotal evidence. I shall use this opportunity to think aloud, albeit retrospectively, about some of my own scientific work, and to see whether it, too, fits the problem-solving mold.

I say "albeit retrospectively" but backward predictions are really the only ones we can trust when we are dealing with a theory of human behavior. After all, when we make forward predictions, our scientists may have been influenced by the very theories of discovery we are trying to test. The theory may fit their behavior only because they have read about BACON or DALTON, and think they will do better science if they simulate those programs. We will avoid that danger of spurious verification by predicting events from a time when the theory did not exist.

Formulating Problems

It is usually thought that a prerequisite to answering a question is to state it. Or, to change the metaphor, for something to be found, something must have been lost. But is that always true? When one finds a vein of gold, was it Nature who lost it? If we can find gold we haven't lost, perhaps we can answer questions we haven't asked.

Let's try again. We may find gold (even gold we haven't lost) by searching for it. But that means that the question has already been asked: "Where can we find *some gold*?" But what about the gold we find when we are not looking for gold, when we are engaged in some quite different activity (gathering wildflowers on the mountain, say)? At the very least, we must *notice* the gold: it must attract our attention, distracting us from the flowers. Do we account for this by postulating a need for gold? Or will an attention-attracting propensity of shiny yellow objects do the job? And how is the attraction of these yellow objects enabled by our distractability from the flower-gathering task?

Now let's *return* from gold-seeking to problem-seeking. If we take our metaphor seriously, it suggests that *one way to find* a problem, and perhaps even its solution, is to try to solve some other problem. That doesn't tell us where the other

problem came from, but one problem at a time! We are dealing with the phenomenon of surprise. Searching for wildflowers, we are surprised to see something shining and golden in the rocks. To be surprised we must attend to the surprising phenomenon. Hence the dictum of Pasteur: "Accidents happen to the prepared mind." And now we have a new problem. How does a mind become prepared? Perhaps it is time for the anecdote.

My first piece of scientific work, begun as a paper in an "independent projects" course at the University of Chicago in the Winter and Spring of 1935, was to study the administration of public recreation in the City of Milwaukee (Simon, 1935). Never mind why that was a problem: it was relevant to a research project of my professor, Jerry Kerwin, on the relations of school boards with city governments. A standard topic in studies of organizations is the budget process, which in this case involved the division of funds between playground maintenance, administered by one organizational unit, and playground activity leadership, administered by another. How was this division (which was a frequent subject of dispute) arrived at?

My previous study of economics provided a ready hypothesis: divide the funds so that the next dollar spent for physical maintenance would produce the same return as the next dollar spent for leaders' salaries. I saw no evidence that anyone was viewing the decision in this way. Was I surprised? Perhaps, initially, but on reflection, I didn't see how it could be done. How were the values of better activity leadership to be weighed against the values of more attractive and better-maintained neighborhood playgrounds?

Now I had a new research problem: how do human beings reason when the conditions for rationality postulated by the model of neoclassical economics are not met -- for example, when no one can define the appropriate utility function, or suggest how the contribution of expenditures to utility is to be measured?

investigating further the particular situation before me. I thought I could see a rather simple pattern of the mental processes. Those who were organizationally responsible for playground supervision wanted more money spent for leadership; those who were responsible for the physical condition of the playgrounds wanted more spent for maintenance. Generalizing, people in organizations bring decision problems within reasonable bounds by identifying with the partial (and more nearly operational) goals that are the particular responsibility of their own organizational units (Simon, 1947, ch. 10).

Of course this is only a partial answer. It defined and labeled the phenomenon of organizational identification, a concept that has proved valuable in *administrative theory, but it did not explain how higher levels of the organization* adjudicated between the claims arising from competing identifications at the lower levels. That subject has subsequently been addressed by other researchers, among *them John P. Crecine*, who wrote his dissertation on this topic some thirty years after the events I am describing (Crecine, 1969).

The broader question -- how do people make decisions when the conditions for the economists' global rationality are not met (or even when they are)? -- remains an active frontier of research today, although large pieces of an answer have been provided through research by cognitive scientists on problem solving. Here the central concept is what economists call "bounded rationality," and what cognitive scientists would more likely label "computational constraints on human thinking." A large part of the answer is that, when people don't know how to optimize, they may very well be able to satisfice -- to find good-enough solutions. And good-enough solutions can often be found by heuristic search (Simon, 1955, 1982).

Now what does this anecdote say about finding problems as an essential component in the process of scientific discovery? One thing it says is that a

problem I found in 1935 has lasted me for fifty two years I have never had to find another. More accurately, this very broad problem of accounting for human rationality has served as a powerful generator for an endless series of subproblems (e.g., how do people solve the Tower of Hanoi problem, how do they choose chess moves, how do they make scientific discoveries?) (Newell and Simon, 1972; Simon, 1979, sections 4, 7; Langley et al., 1987).

Another lesson to be drawn from the anecdote is that scientific discovery is incremental. An explanation for a particular act of discovery must take everything that has gone before as initial conditions. What we seek to explain is how these initial conditions led to the next step -- in this case how my knowledge of elementary price theory, and Jerry Kerwin's desire to know how the school board and the public works department cooperated to provide public recreation services in Milwaukee, led me to observe a phenomenon that initially surprised me; and how that surprise led to new observations that could be explained by the concepts of identification and bounded rationality. Steps taken twenty years later led from bounded rationality to satisficing, and from satisficing to heuristic search.

Third, the anecdote adds another to the long list of examples where surprise was a key element in discovery. But what was "prepared" about this particular mind? My training in economics, and the evocation of that training in the context of a budget situation, disclosed a contradiction between what theory taught me ought to be happening and what my eyes and ears showed me to be actually happening. Without the training in economics the observed behavior would have appeared entirely "natural." Without the observations, I could have continued in the happy illusion that the neoclassical theory of utility maximization explains human behavior in the domain of budgeting.

Nothing mystical. Nothing magical. Can we simulate it? The heuristics indeed

resemble quite closely those of KEKADA, the program that Deepak Kulkarni and I have used to simulate the research strategy of Hans Krebs, who found the chemical path for the *in vivo* synthesis of urea, a program that has now been generalized to other discoveries (Kulkarni and Simon, 1986). The program experiences surprise when its expectations are not met and reacts to its surprise by seeking explanations for the surprising phenomena. But now I am waving my hands. (Or am I hand simulating?) We have not yet investigated what heuristics would have to be added to KEKADA in order to simulate the discovery of bounded rationality. But I think I might have saved myself a lot of work in 1935 if I had had KEKADA to advise me.

Laws from Data

In our book, *Scientific Discovery* (Langley et al., 1987), my colleagues and I gave primary attention to the problems of inducing generalizations, quantitative and qualitative, *from* empirical data. Our programs, BACON and DALTON, were systems of heuristics for *inducing* quantitative laws, and our programs STAHL and GLAUBER systems for inducing qualitative laws.

Data are not the only possible initial conditions for the induction of new laws; theories can also be used, in conjunction with data or independently. In our BACON simulations we showed that by incorporating in BACON heuristics that search for symmetries and conservation laws, we could substantially improve the efficiency with which it found laws in empirical data. In the limit, it may be possible to find a descriptive law directly, by deriving it from a more fundamental explanatory law. For example, Newton showed that Kepler's Third Law of Planetary Motion (the period of revolution of a planet varies as the 3/2 power of its distance from the Sun) could be derived mathematically from the inverse square law of gravitational attraction. (But note that Newton was working backward from the law that Kepler had already

discovered by data-driven search.)

Before one can find mathematical functions that fit empirical data, one must have appropriate data that look as though a smooth mathematical function could generate them. It's the recipe for rabbit stew all over again: first catch the rabbit. Examples of such data have been much easier to come by in the physical sciences than in the biological or social sciences. When we find social science data of this kind, we should prize them.

On only one occasion in my life have I run on to such data, and I cannot recall exactly when I first encountered them -- possibly as early as about 1936 in Lotka's *Elements of Physical Biology* (1924), a fascinating book that the economist Henry Schultz always called to the attention of his students. Lotka reports data, compiled by one Dr. J.C. Willis, showing that when the number of species belonging to each genus in some order of plants or animals (beetles, say) are counted, and the genera are then arranged in order, according to the number of their species, the genus with the nth largest number of species will have about 1/n as many species as the genus with the largest number.

Similarly, when the frequencies with which different words appear in a book are counted, and words are then arranged in order of *their frequency*, the nth most frequent word will occur about 1/n times as frequently as the most frequent word. Moreover, about half of all the words that occur in a book will occur exactly once, about one sixth exactly twice, one twelfth three times, and so on. These relations hold for books in any alphabetic language, and the departures *from* regularity are small.

Other data show a similar regularity in the populations of cities in the United States: the nth largest city is about 1/n times as large as New York. These regularities are easily seen if the data are plotted on log-log paper, whereupon they

fall on a straight line with a slope of minus one

What does one do with regularities like this -- regularities that at first blush can only be described as astonishing? What one does (or should do) is to behave in a BACON-like fashion until one finds a formula that fits the data Then like DALTON one should see if one can postulate a mechanism whose operation would produce the regularity described by the formula

I wish I could say that this was my immediate response to the data Memory fails me I recall my fascination with them but not whether I pondered over them, and if I did, for how long I do recall that, when I returned to Chicago after 1942, I thought about them again -- I have a clear picture of sitting in the Biology Library in the University of Chicago reading a paper referenced in Lotka's book. I also recall mentioning my interest in the matter to Allen Newell while visiting him and Noel in their Santa Monica apartment between 1952 and 1954. But I was doing many other things during these years The startling data on word frequencies and city sizes were not a constant preoccupation, but were more like a recurring itch that needed to be scratched occasionally

Sometime during 1954 I found the answer My recollections of just how I found it are sketchy, with no scraps of paper to bolster my memory, but a few aspects of the discovery are recoverable now First, I looked for a function to fit the data I was especially impressed by the regularity of the word-frequency data at the low end of the frequency range The simple fractions seemed to point to a formula involving ratios of integers In fact the simple formula $f(i) = 1/[i(i+1)]$, gives the required numbers, 1/2, 1/6, 1/12, and so on For large i, we have approximately $f(i) = 1/i^2$ The rank, which is simply the integral of the frequency, will then give $F(i) = 1/i$ so that on a logarithmic scale the relation between rank and frequency will be linear with a slope of minus one

Finding an equation that fits these magic numbers sets the stage for a new problem: finding an explanation for the equation, a plausible mechanism that will provide a rationale for the phenomena. My recollections of how I did this are even sketchier than my recollections about the previous stage. The ratios of integers were again the key. Where can you get ratios of integers? Ratios of factorials are one possible source. 1/6 can be written as the product of 1/2 and 1/3, and 1/12 as the product of 1/2, 1/3, and 2/4. In general, the formula (n-1)!/n! produces the required numbers. The next step is likely to occur only to someone who has a little mathematical knowledge, and who sees in these ratios of factorials something like the Beta function, or at least sees the kinds of expressions one is accustomed to encounter in problems on combinations and probabilities. (In fact, I discovered that the Beta function was what I wanted by searching through my copy of Peirce's *A Short Table of Integrals*, where I vaguely remembered having seen some ratios of factorials.)

Are there any other reasons for thinking that the situation may call for a probability model? Indeed there are. What do word frequency distributions and city size distributions (as well as a number of quite different phenomena where this same law applies) have in common? Nothing very obvious, unless they can be viewed as instantiations of the same urn scheme. So *let us see whether* we can interpret the formula as *representing the steady state of some* stochastic process.

Here I recall being aided by a metaphor. If we think of a book as being created word by word, and if a word is added that has already occured k times, the number of words occuring k+1 times each will be increased by one, and the number of words occuring k times each decreased by one. For a steady-state equilibrium the rate at which words are created that had previously occured k times must be equal to the rate at which words are created that had previously occured k-1 times

In this way the "k" bin will be replenished as rapidly as it is depleted At some point I began to visualize this as a cascade, with successive pools of water each maintained at a constant level by flow in from the pool above, and flow out to the pool next below Working back from our answer -- the distribution that we know describes the phenomena -- it is not too hard to show that the equilibrium condition requires that the probability of creating a word that has already occured k times must be proportional to k.

We are ready for the final step to interpret the probability assumption. In the case of word distributions, it can be interpreted to mean that the chance of a word being chosen as the next word in a text is proportional, because of association, to the frequency with which it has been used already, and also proportional, because of long-term associations stored in memory, to the frequency with which it is used in the language. In the case of city sizes, it can be interpreted to mean that birth and death rates are approximately independent of city size, while the probability that a city will be the target for any given migration is also proportional to its size (Simon, 1955).

I don't propose to defend these interpretations here My purpose is to understand the process that reached them If my account, through the filter of thirty to fifty years of forgetting, has any relation to reality, then we see a process for arriving at the initial formula that looks very BACON-like, followed by working-backward search processes that are guided by the evocation of prestored mathematical and real-world knowledge -- BACON as the front end to an expert system

Again, my hands are waving wildly. You will not have failed to notice that I have not accounted at all for the cascade metaphor, yet at some time it was evoked and helped me to formulate the steady-state relations So there is still work to be done on the theory of discovery, still theses to be written and papers published But

I see in this little history, or imagined history, no magic and no mystery. Each step appears to proceed, if not inexorably at least plausibly, from the preceding one.

If the data cried out so loudly for explanation, and if the discovery process proceeded in such a plausible succession of steps, why did not others discover this law and its stochastic explanation? Indeed they did. The first was G. Udny Yule, the English statistician, who in 1924 published "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S." Yule constructed a model very similar to the one I have just described to explain Willis' data, mentioned earlier, on the distribution of species among genera. (I could have been led to this paper by a footnote in Lotka, but I wasn't.) A second was the English economist, D. G. Champernowne, who published "A model of income distribution" in 1953, describing a quite similar process. A third was B. Mandelbrot, who, in 1953, published "An informational theory of the statistical structure of language." There were some differences between Mandelbrot's model and mine, which later occasioned heated dispute between us, but the basic ideas were closely related.

I learned about all of these partial anticipations when I searched the literature and inquired among my friends prior to publishing, in 1955, my own paper on the topic (Simon, 1955).

That still isn't quite the end of the story, for again, the solution of one scientific problem created a host of new problems. In the book by Yuji Ijiri and myself, *Skew Distributions and the Sizes of Business Firms* (1977), you can find a series of essays applying the same stochastic mechanism, and generalized versions of it, to the task of understanding the size distributions of business firms and the economic implications of these distributions.

Representations

Mention of the cascade metaphor that I used to find the stochastic process underlying word-frequency and city-size distributions raises the question of representations. What kinds of representations are used by scientists in thinking about their research problems, and where do these representations come from? One hallowed form of the question is whether scientists (and others) think in words, or whether thoughts take some quite different shape -- whether they employ "mental pictures," say.

Words and Pictures

The French mathematician, Jacques Hadamard, in his delightful essay on *The Psychology of Invention in the Mathematical Field* (1945), comes down heavily on the side of images and against words. Among the many distinguished mathematicians and scientists testifying for him is Albert Einstein, who in a letter to Hadamard stated that "the words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be 'voluntarily' reproduced and combined."

What is good enough for Hadamard and Einstein is good enough for me. I too, have difficulty in finding any presence of words when I am thinking about difficult matters, especially mathematical ones. Even as I sit here at the keyboard, composing this essay, I cannot really detect the words in my thoughts (or much of anything else, for that matter) until they come out the ends of my fingers. But perhaps I am not thinking, but just recording previously composed ideas that reside somewhere in my subconscious mind.

Even if we do think in images rather than words, neither Hadamard nor

Einstein nor I had much success in describing just what these images were or how they were represented in a biological structure like the brain. However, I believe that Jill Larkin and I have recently made substantial progress in explaining these matters (in *Why a Diagram is (Sometimes) Worth Ten Thousand Words*) (Larkin and Simon, 1987). In order to deal with the difficulties one by one, we fudged a bit, alleging that we were talking about diagrams on paper rather than mental pictures, but most of our argument carries over in a straightforward way. The basic ideas, which I will not elaborate upon here, are (a) that in the course of transforming verbal propositions into images, many things are made explicit that were previously implicit and hidden, and (b) that (learned) inference operators facilitate making additional inferences from the images in computationally efficient ways.

We also show, as a byproduct of our analysis, that diagrams are representable as list structures (alias "schemas," "scripts," "frames," "labeled directed graphs," et cetera), hence are programmable in standard list-processing languages, hence are readily seen to be representable in systems of neuron-like structures. Since the surface structures and the semantics of natural languages can also be represented as list structures, we can conclude that propositions and pictures (or at least diagrams) can use common machinery in the brain -- that both are best viewed as specializations of a common list-structure mode of representation. (I do not wish to deny that we may *also*, as Kosslyn (1980, and also this volume) has argued, possess a specialized raster-like organ for more literal representation of visual images -- a *mind's eye*. But I would prefer to put that question aside here.)

Now just as there has long been a debate as to whether we use words or images in our thoughts, so there has been a debate (perhaps the same debate) as to whether our internal representations of problems look like collections of propositions or like models of the problem situations. Each of these views has been

held by an important segment of the cognitive science community, and the two segments do not often communicate with each other except sometimes to quarrel.

One segment, under the banner of "let language lead the way," takes verbal reasoning as its metaphor for the problem-solving process, and thinks of reasoning as some kind of PROLOG-like theorem-proving procedure. The book by Miller and Johnson-Laird on *Perception and Language* (1976) is an excellent representative of this point of view, although Johnson-Laird, in his more recent book, *Mental Models* (1987), takes a long step of apostacy toward the alternative viewpoint. That he does so *without any apparent* awareness that he is moving onto well-explored ground exemplifies the mutual insulation of the two segments.

The second segment of the cognitive science community uses heuristic search through a problem space (a mental model of the task domain) as its metaphor for problem solving. *Human Problem Solving* (1972) adheres strictly to this viewpoint. It has been claimed, by Pylyshyn (1973) among others, that the two viewpoints cannot be distinguished operationally, but this claim rests on a confusion between the informational equivalence and the computational equivalence of representations. Even if two representations contain exactly the same information, it may be far cheaper, computationally, to make some of this information explicit using one rerpresentation than using the other. The incorrectness of the claim of computational equivalence is demonstrated by the examples given in the Larkin-Simon paper mentioned above.

Representing a Dynamic System

I am afraid that I have been diverted from my main topic, which is providing anecdotal evidence about the problem solving processes used in scientific discovery. Let me return with an example that I will present rather sketchily, to avoid technical detail. Economists frequently use what they call "partial equilibrium analysis," in which they avoid talking about everything at once by making a host of *ceteris paribus*

assumptions They examine the impact of a disturbance upon a small segment of the economy while assuming no interaction with the rest of the economy

If challenged on the legitimacy of what they are doing, economists using partial equilibrium methods may defend themselves by saying that, of course, interactions are not completely absent but they are small, hence unimportant That is an argument we know not only in economics, but throughout all of science But is it a satisfactory argument? Small effects, persisting over a long period of time, may integrate into large effects

Thoughts of these kinds (represented as words or as images?) went through my mind while I read, in the early 1950s, a paper by Richard Goodwin, *Dynamic Coupling with Especial Reference to Markets Having Production Lags*, published in *Econometrica* in 1947. Again, I cannot claim any clear recollection of the precise steps I took to formulate and solve the problem that his paper evoked. I did conceive of it as a matter of analysing the behavior of a large dynamic system divided into sectors, with strong interactions among the components in each sector and weak interactions among sectors. I remember also that I worked very hard for several months to get answers, and that I worked, without paper and pencil, while taking long walks

My representation, at least much of the time, was an image of the matrix of coefficients of such a dynamic system -- hardly surprising, since this is the way dynamic systems are normally represented in mathematics books At some point I saw that the rows and columns of the matrix could be permuted so that the new matrix would consist of a number of diagonal blocks with large coefficients in them, and only small coefficients in the matrix outside the diagonal blocks The matrix was "nearly block diagonal" The image was vague, in that the number of blocks and their sizes were not seen in detail If forced to give numbers, I might say that there

could have been three blocks, each three rows by three columns in size -- but the answer is surely a *fabrication*.

At some later point in time, I acquired a metaphor. I visualized a building divided into rooms, each of which was divided, in turn, into cubicles. You can see a diagrammatic interpretation of my metaphor on page 212 of *The Sciences of the Artificial* (1981), Second Edition. We start out with an extreme disequilibrium of temperature, each cubic foot of each cubicle being at a different temperature from its neighbors.

Several things now seemed obvious. Throughout each cubicle, a constant temperature would be established very rapidly by exchange of heat between adjoining volumes. At some later time, each room would attain a constant temperature by heat diffusion through the walls of the cubicles. At a still later time, the entire building would reach a constant temperature by exchange of heat between the thicker walls of the rooms.

Moreover, because of the differences in the durations involved, each of these processes of equilibration -- within cubicles, among cubicles, and among rooms -- can be studied independently of the others. In studying the equilibration of each cubicle, we can ignore the other cubicles. In studying the equilibration of rooms, we can represent each cubicle by its average temperature, and ignore the other rooms. In studying the equilibration of the building, we can represent each room by its average temperature. As a result, the mathematics of the problem can be simplified drastically.

There still were some difficult mathematical steps from this picture of the situation to rigorous proofs of the (approximate) validity of the simplification, but the result to be attained was clear. The reasoning I have described was carried out mainly in the summer of 1956, and incorporated, together with the mathematics, in a

paper written with Albert Ando later that year, and published in 1961 (Simon and Ando, 1961).

I can throw no further light on the source of the heat-exchange metaphor, or on how, if at all, I drew inferences from the image of the nearly block diagonal matrix. Block diagonal matrices were not unfamiliar to me, for they had played an important role in the theoretical work I had done on causal ordering in 1952 and 1953. The mathematics required for the proofs, which was fairly standard, would have been evoked, I think, in the mind of any mathematician who had put the problem in the form we did. Our results were rediscovered by some Russian mathematicians in the late 1960's, but apparently had not been anticipated earlier (Korolyuk, Polischuk and Tomusyak, A.A.,1969).

Our theorems and methods (which may be used to invert matrices that are nearly block diagonal) have attracted the attention of numerical analysts, and of natural scientists who are concerned with hierarchically organized systems. The aggregation method we introduced has also now been recognized to be closely related to the so-called "renormalization" procedures that play an important role in several parts of physics, and which were also invented quite independently of ours.

Even with this sketchy account, the discovery process appears quite unremarkable. The problem was found in the literature (Goodwin's paper), and it can be represented in a quite standard way by matrices having a certain special structure. The metaphor, by showing how such a system would behave, made clear the nature of the theorems to be proved. Although nothing is revealed about the source of the metaphor, it is not at all esoteric. The proofs, while intricate, would not pose any great difficulty for a professional mathematician. A case of normal problem solving, we would have to conclude

Finding an Explanatory Model

The last two sections provided two examples of the process of finding an explanatory model -- a model for the rank-frequency relation and a model of nearly decomposable dynamic systems. The reader may be interested however in an example a little closer to home. How could one discover an explanatory model of human problem solving? One answer might be "By observing *some problem solving* behavior closely and inducing the model directly from your observations."

There is a good deal of merit in that answer, and in a later section I will show that something like that happened when the General Problem Solver was invented. But even in this case, the empirical observations were not the sole source of information that guided the discovery. The inventors also had some notions of the shape of the thing they were looking for.

Explanatory theories take a variety of forms. For example, the behavior of gases is commonly explained by supposing that they consist of a cloud of energetic particles, interacting with each other in accordance with the laws of mechanics. Magnetic attraction between two bodies is explained by a *field of magnetic* force in the space between them.

One very common form of explanation, in both natural and social science, employs systems of differential equations or difference equations to determine the values of the time derivatives of system variables. At any given time, the system is supposed to be in a specified "state," and the differential equations then determine to what state it will be moved a "moment" later. Thus, in mechanics, the state is defined in terms of positions and velocities, and the differential equations show how the action of forces to produce accelerations brings about a continuing change in state through time.

Building an explanatory model involves a choice among *these or other*

representations of the phenomena. Will it be a particle model or a continuum model? Will it represent static equilibrium, a steady state or dynamic change? The representation has to be chosen prior to, or simultaneously with, the induction of the model from the data.

When Allen Newell, Cliff Shaw, and I began the construction of a theory to explain problem solving, around 1955, we were already committed to a representation. In fact, it was our recognition that such a representation had become available with the invention of the digital computer that motivated us to undertake the study of human thinking. A reader can find detailed accounts of the background for this recognition in Newell and Simon, 1972, pages 873-886, and in McCorduck, especially Chapters 3 and 6.

What we observed -- we have told the story before (Newell and Simon, 1972) -- was that the program of a computer is formally equivalent to a set of difference equations. At each operation cycle, the program determines the new state of the machine as a function of its previous state (the contents of all its memories) together with any new input it has received. Moreover, these difference equations were not limited to manipulating numbers, but could process symbols of any kind.

The explanatory task, then, was to find a dynamic theory of the processes of problem solving in the form of a computer program. The data we could muster on the behavior of human problem solvers had to be examined for clues as to the nature of that program. This requirement provided very strong guidelines both for the kinds of data that would be valuable (preferably data that followed the course of problem solution as closely and minutely as possible), and for the best ways of examining the data (searching out the succession of "actions" the problem solver executed, and the cues that motivated each action).

Of course, there was more to the representation than simply the specification

that it be a computer program It had to contain symbol structures that could represent the structures in human memory which were known to be in some sense associative The important point was that there was a continuing two-way interaction between the gradual construction of the representation and the construction of the theory that used it Sometimes programming convenience (or necessity) dictated choices sometimes psychological requirements Some aspects of the representation that were initially conceived mainly to meet programming needs (for example, the list-processing languages and data structures in the form of lists and description lists), were later seen to have psychological interpretations as networks of associations.

The empirical part of the undertaking which I will discuss later in connection with the topic of experimental design went hand in hand with the design of the representation of the explanatory model

Once some experience had been gained with information processing models in the form of computer programs, they became a readily available tool for buiding theories of other aspects of human thinking So Kotovsky and I, interested in explaining simple law-discovery processes as a first step toward a theory of scientific discovery "naturally" framed our model as a computer program in a list processing language, capable of discovering and extrapolating the patterns in Thurstone letter-sequence problems (Kotovsk, and Simon in Simon 1979 ch 5 1, 5 2) No alternative representations were even considered

In the past few years with the availability of a whole new menu of variants production systems models of memory with spreading activation connexionist mode's SOAR the PROLOG language -- choices of representation have again become an important and difficult part of the model-building process

Designing Good Experiments

Experiments are supposed to be aimed at testing hypotheses or better yet choosing between contending hypotheses ("critical" experiments). That an experiment meets one or both of these aims is neither a necessary nor a sufficient condition for its being a good experiment.

It is not a sufficient condition because testing weak-tea hypotheses of the form "variable X affects variable Y." or its negation, is not usually very interesting, and does not often contribute much to our understanding of the world. (But if I continue in this vein, I will trespass on Allen Newell's noted "Twenty Questions" essay (Newell 1973).)

Testing stronger quantitative hypotheses (e.g., the periods of the planets are as the 3/2 power of their distances from the Sun) is much more interesting, and very interesting indeed if the hypotheses are closely connected with broad explanatory theories (e.g., with the inverse square law of gravitation).

We are on safer ground if we aim experiments at testing *models* instead of testing *hypotheses*, but when we do that, we must remember to throw away the whole standard apparatus of significance tests, which is no longer applicable. (See Gregg and Simon, and the references cited there (in Simon, 1979, ch. 5.4).) We must also remember that models are multi-component creatures, and when our data don't fit a model, we are faced with a difficult diagnostic task to determine what to change -- or whether to discard the entire model.

So much for sufficiency, what about necessity? Is model-testing the only reason for experimenting? Surely not. One good reason for running an experiment -- or for spending one's time just observing phenomena closely -- is that you may be surprised. The best things that come out of experiments are things that we didn't expect to come out -- especially those that we couldn't even have imagined in

advance as possibilities. Of such stuff are many Nobel Prizes made.

Lest I be accused of advocating planning experiments by casting dice, let me suggest that there are heuristics for planning both kinds of experiments: experiments to test models and experiments to generate surprise, and let me illustrate the heuristics with some examples. I will begin with the more traditional model-testing category.

Testing Models

A few years ago, I found occasion to begin the study of the Chinese language. I did it just for fun, and because I planned to visit China, but to put a more solemn face on things, I called it "exposing myself to new phenomena." That allowed me to do some of it on company time, with a good conscience. Finding myself in China, working with Chinese psychologists, we decided to replicate with Chinese language materials some standard short-term memory experiments. The motive was to test a model. Does Chinese have a magical number? And is it seven? The answer to both questions was "yes" -- no great surprise.

Meanwhile, I had learned a striking fact about the Chinese language (no surprise to my Chinese colleagues, but a surprise to me). A Chinese college graduate can recognize about 7,000 Chinese characters (hanzi). Each character is pronounced with a single syllable. But in the Chinese language there are only about 1,200 distinct syllables (even taking account of tone distinctions). Hence, on average, there are about six homophones for each character.

Somehow (intuition or recognition at work), I remembered that short-term memory was generally thought to be acoustical in modality, but only because of Conrad's rather indirect evidence that errors in recall generally involved similarity in sound rather than similarity in appearance. In Chinese, we could put the acoustical hypothesis to direct test. After establishing that the STM span is about six or seven

unrelated and non-homophonic characters we presented the same subjects with strings of visually distinct homophonic characters The result was dramatic -- the STM spanned dropped to about two confirming Conrad s result (Zhang and Simon 1985. Yu. Bolin et al 1985)

Similar methodological predispositions underlie the experiments that Bill Chase and I did on memory for chess positions building on the earlier work of De Groot and others (Chase and Simon in Simon 1979 ch 6 4 6 5) Here the question was whether the differences in chess memory between experts and novices could be accounted for by differences in their vocabularies of "chunked" chess patterns The answer that came out of our experiments was "qualitatively yes. but quantitatively no." an answer that. if slightly disappointing. was much sharper than if we had simply asked whether experts' chunks were larger than novices'

The qualified affirmative answer has led to much subsequent research which is gradually giving us a more precise model of how chunks are constructed and organized in memory. This research is strong represented in this conference by the paper of Charness. and that by Ericsson and Staszewski. while closely related questions are examined by Carpenter and Just in their use of eye movements and a memory model to study the role of working memory in reading

The experiments with Chase on chess memory. like those with Chinese characters. were designed by asking what quantitative predictions a current model made and how we could make the measurements necessary for testing these predictions The problem solving search. if there was one took place in the space of the characteristics of the task domain. and was facilitated by looking for "surprising" or "interesting" features of the domain In the Chinese language case. the surprising feature was found first. and the model it was relevant to was found second In the chess case. the order was reversed

The experiments just described all have an experimental and a control condition, just as a well-designed experiment is supposed to. In the Chinese language experiments we compared homophonic with non-homophonic strings of characters. In the chess experiments, we compared the performance of experts with the performance of novices, and chess positions from well-played games with random positions. The expert-novice dichotomy has also served me in good stead in some more recent experiments on problem solving in physics (Simon and Simon, 1978; Larkin et al., 1980), and has been used by Hayes (this volume) and his colleagues in research on writing. An incidental benefit of using this paradigm is that being able to point to clearcut experimental and control conditions seems to soothe the savage breast of referee and editor.

Problem Isomorphs

To conclude my list of experimental manipulations, I will mention just one other, which has provided us with almost unlimited mileage -- the idea of problem isomorphs. Its history is as obscure as that of many of the other things I have been talking about. I think I invented the idea of problem isomorphs about 1969, or a little earlier; for I do not have any evidence of earlier mention by myself or anyone else. I have a conjecture about its antecedents, but it is a reconstruction, not a recollection, although John Michon, without prompting, corroborated it.

Saul Amarel was one of the first researchers in artificial intelligence to point out that changing the representation of a problem -- the problem space and operators -- could sometimes greatly facilitate its solution. Amarel, Newell, and I participated in a semester-long seminar at CMU in 1966, the main topic of which was problem representation. Now it is only a small step (at least by hindsight) from the idea that a subject can solve a problem easily by finding the right representation to the idea that an experimenter can make a problem harder or easier for a subject by

presenting it in one or another guise

So much for the antecedents Soon problem isomorphs -- problems with identical task domains and legal-move operators, but described by different sets of words -- were a topic of discussion in the Understand Seminar (alias the Cognitive Science Seminar) which has run weekly in the Psychology Department for twenty years The first example was number scrabble, an isomorph of tic-tac-toe, and John Michon then added another member, JAM, to this set. John R Hayes rapidly became the most prolific and ingenious designer of problem isomorphs, providing us with somewhere between a dozen and two dozen isomorphs of the Tower of Hanoi puzzle, most of which have been used in one or more experiments (Hayes and Simon, in Simon, 1979, ch 7 1-7 3).

We have used isomorphs to discover what characteristics of a problem, other than the size of the task domain, account for its difficulty. Early work in problem solving, our own included, had focused on the combinatorial explosion of search as the main source of problem difficulty Yet we had found that the Tower of Hanoi, with a relatively small and easily exhaustible domain, and the Missionaries and Cannibals puzzle, with a tiny one, could occupy human adults for fifteen minutes or a half hour before they found a solution.

The idea that only the size of the task domain could affect problem difficulty sometimes died hard. One referee for a funding agency gave a project proposal low marks because he or she thought that, on these grounds our experiments could have only negative results, since all isomorphs must be of the same difficulty (At the time we were told of this objection, we had already demonstrated differences in average solution times to various isomorphs of the Tower of Hanoi in the ratio of 16 to 1) Our theoretical account of problem difficulty is still very incomplete, although we have been able to model some of the phenomena (Kotovsky, Hayes and Simon

1985. Kotovsky and Fallside this volume) (The UNDERSTAND program, for example, constructs different representations of problems when presented with different isomorphs (Hayes and Simon, in Simon, 1979, ch. 7.1).)

The experimental strategy here is clear deriving from the single idea of isomorphism -- and that idea has at least a plausible lineage. The power of the idea would be enhanced if we had more systematic ways of designing isomorphs with *specific* features designed in advance to test particular putative sources of difficulty.

Experimenting Without an Independent Variable

The experiments described up to this point all compare performance under two or *more different* conditions -- they all involve manipulation of an independent variable. When I examine my publications beyond the limited set already mentioned, I find to my embarrassment that this fundamental condition for sound experimentation is seldom met in them. What have I been up to? What can I possibly have learned from ill-designed experiments? The answer (it surprised me) is that you can test theoretical models without contrasting an experimental with a control condition. And apart from testing models, you can often make *surprising* observations that give you ideas for new or improved models.

Let me start with an example of the latter kind. Many summers ago (about 1965) Jeffery Paige and I decided to take thinking-aloud protocols from high-school students solving algebra word problems (Paige and Simon, in Simon, 1979, ch. 4.4). Our main motivation, I think, was just to see how they did it -- what processes they used. Perhaps we had in mind comparing their behavior with Bobrow's STUDENT program (Bobrow, 1968), which solved such problems. Or perhaps we thought we might build a program ourselves that would do a better job of simulating the human processes. If those were our intentions, my memory does not retain them.

Jeff conceived of a fine idea (at least, I have always remembered it as his)

We constructed some "impossible" problems -- problems that could not be given a real physical interpretation because their solutions involved boards of negative length or nickels that were worth more than dimes We then asked our subjects to set up the equations corresponding to the problem statements, but not to solve them

The outcome was wholly unanticipated Our subjects fell into three groups, rather consistently over the set of three problems. Some set up the equations that corresponded literally with the verbal statements of the problems. Some translated the problems inaccurately, always ending up with equations that corresponded to a realizeable physical situation. Some said, "Isn't there a contradiction?" -- meaning, "I draw inferences from the problem statements that conflict with my knowledge of the real world."

Because we were trying to get as dense a set of data as we could, in order to trace processes in detail, we had asked the subjects both to think aloud and to draw diagrams of the problem situations. The diagrams drawn by subjects in the first group were generally incomplete and unintegrated, and did not reveal the "contradiction." The diagrams drawn by subjects in the second group misrepresented the situations in just the way their equations did -- so as to make them physically realizeable The subjects in the third group drew diagrams that revealed the contradictions

The direction of the causal arrow is not clear, but one can take these results as at least presumptive evidence that subjects in the second and third groups used imagery to represent the information from the word problems before translating into the language of algebra Subjects in the first group gave evidence of translating directly to equations using only syntactic information.

With this kind of information in hand, one can begin to construct models for the simulation of these sorts of behavior, and to explore what other predictions could

be made about systems behaving in these ways. The ISAAC program, written by Gordon Novak to solve physics problems presented in natural language, is an example of a system that uses an internal diagram of the problem situation to mediate between the verbal stimulus and the equations it finally constructs (Novak, 1976). The UNDERSTAND program that John R. Hayes and I constructed, around 1972, to show how verbal problem instructions could be converted into inputs appropriate for a GPS-like problem solver, also borrowed this insight from the algebra experiments (Hayes and Simon, in Simon, 1979, ch. 7.1).

But the most massive set of examples of the experimental strategy of "just looking" is to be found in *Human Problem Solving.*" Density of data was the name of the game, and protocol analysis the way of playing it. In 1956, the *Logic Theorist* (LT) had demonstrated the feasibility of solving difficult problems by highly selective heuristic search (Newell and Simon, 1956). But is that the way people did it? The *General Problem Solver*, LT's successor, was our answer -- a heuristic search system that used means-ends analysis as its principal heuristic (Newell and Simon, 1972).

Both Al Newell and I agree that the core of GPS was extracted *directly* from a particular protocol that we can identify. We also agree as to the week in the summer of 1957 when it was done. On the details, the evidence is not wholly concordant, but sometime, when we have leisure to examine the papers we have preserved, we may get it all straightened out (See McCorduck, page 212). The main lesson is clear: GPS, a theory of human problem solving, was extracted by direct induction from the thinking aloud protocol of a laboratory subject, without benefit of an "experimental" and a "control" condition.

What, in addition to luck, entered into the result? First, as I have pointed out in an earlier section, we already knew that we *wished to represent* our model as a computer program in a list-processing language. Second, a data gathering method

was used that obtained the densest record of the subject's behavior that we knew how to get. We were able to discover what he had done each few seconds of time during which he worked on the task. Third, some care had been taken in selecting the task. It had already been used by O. K. Moore and his colleagues at Yale, and we had access to both their experience and their data. The task was symbolic, hence made for easy verbalization, and seemed to call for a minimum of pictorial visualization. It was a hard enough task to evoke genuine problem solving behavior from intelligent subjects.

Application of these criteria to the selection of problem solving tasks accounts for a substantial fraction of the knowledge that has been collected about problem solving processes during the past thirty years, and a substantial part of the theoretical efforts that have succeeded in building models to account for behavior in many kinds of tasks. The metaphor of chess, cryptarithmetic and the Tower of Hanoi serving as the green peas, Drosophila and E. coli of cognitive science is as near to literal truth as it is to fancy.

Do these experiments really lack independent variables? Can't we consider the task domain or the subject to be just that? Of course we can, but to no particular end. The principal knowledge we gained from these experiments did not come out of comparisons between tasks or subjects. It came out of painstakingly analyzing individual protocols and inducing from them the processes that problem solvers employed in their work. Once this had been done, we could test the generality of our results by comparing over tasks and over subjects. But detailed longitudinal analysis of the behavior of single subjects was *the foundation* stone for the information processing theories we have built of what goes on in human problem solving.

If the methodology troubles us, it may be comforting to recall that detailed

longitudinal analysis of the behavior of a single solar system was the foundation stone for Kepler's Laws, and ultimately for Newton's. Perhaps it is not our methodology that needs revising so much as the standard textbook methodology which perversely warns us against running an experiment until precise hypotheses have been formulated and experimental and control conditions defined. How do such experiments ever create surprise -- not just the all-too-common surprise of having our hypotheses refuted by facts, but the delight-provoking surprise of encountering a wholly unexpected phenomenon? Perhaps we need to add to the textbooks a chapter, or several chapters, describing how basic scientific discoveries can be made by observing the world intently, in the laboratory or outside it, with controls or without them, heavy with hypotheses or innocent of them.

The Scientist as a Satisficer

My economist friends have long since given up on me, consigning me to psychology or some other distant wasteland. If I cannot accept the true faith of expected utility maximization, it is not the fault of my excellent education in economics -- in fact, the education was repeated four times, often enough even for a slow learner. First, as a high school student, I read the works of Richard Ely and Henry George in order to meet the arguments of opposing debating teams on such issues as the tariff or the single tax. Then, at the University of Chicago, I learned price theory from Henry Simonds and Walrasian equilibrium and econometrics from Henry Schultz.

Next, at Berkeley, my colleagues, Kenneth May and Ronald Shepard, students of Griffith Evans, revealed to me the inference-drawing powers of the second-order conditions of maximization, while I learned about Neyman-Pearson statistics from Jerzy Neyman himself. Finally, on returning to Chicago, I was exposed to Samuelson's

Foundations and Hicks on value in the brilliant discussions at the Cowles Commission seminars among Jascha Marschak, Tjalling Koopmans, Ken Arrow, Larry Klein, Franco Modigliani, Gerard Debreu, and other superbly keen and well-informed minds.

Alas, it did not take. My traumatic exposure in 1935 to the budgeting process in the Milwaukee recreation department had immunized me against the idea that human beings maximize expected utility, and had made of me an incorrigible satisficer. And that same imprinting experience supplied me with the problem -- the cornucopia of problems -- that has kept me occupied ever since. I have sketched here the theory of scientific discovery to which my study of these problems has led me. It is not a theory of global rationality, but a theory of human limited computation in the face of complexity. It views discovery as problem solving, and problem solving as heuristic search, and heuristic search as the only fit activity for a creature of bounded rationality.

Some scientists believe that theories should be judged by their ability to make correct predictions. This paper provides some tests of the predictive power of this problem-solving theory of discovery. The anecdotes I have provided from my own scientific life are instances where it gives a pretty good account of the processes that are visible in my research.

It describes me, like KEKADA, formulating a new problem in response to my surprise at encountering an unexpected phenomenon. It traces my BACON-like progress toward discerning a lawful regularity in data, and the evocation of knowledge, in expert-system style, to find an explanation for the regularity. It accounts for my use of diagrams to gain a grasp of complex phenomena in a dynamic system. It illuminates how the availability of representations and the invention of new ones has influenced my efforts to construct explanations. It characterizes a number of my strategies for designing experiments, and perhaps even explains why I

am frequently unconcerned about such things as "experimental controls" or even independent variables.

Of course I am exercising poetic license in talking of predictions. A comprehensive SIMPLE SIMON has not been programmed, only pieces of him exist. It would be more defensible to talk of explanatory accounts rather than predictions. But you will not be misled by the metaphor, which is as useful as one can expect a metaphor to be.

The information processing theory of discovery that I have been describing has one other virtue. It is not only a descriptive theory, but a normative one as well. Not only does it predict (explain) my behavior successfully, but, unbeknownst to me, it has served me for fifty two years as a reliable set of heuristics for conducting research. Quite unwittingly, I have been following the instructions of BACON, of STAHL, of GLAUBER, of DALTON, and of KEKADA. I couldn't have had better guidance.

However, one heuristic that has been of first importance to my work is missing from these programs. I will mention it, because you too may find it useful. If you want to make interesting scientific discoveries, be sure to acquire as many good friends as possible, who are as energetic, intelligent, and knowledgeable as they can be. Form partnerships with them whenever you can. Then sit back and relax. You will find that all the programs you need are stored in your friends, and will execute productively and creatively as long as you don't interfere too much.

References

Bobrow, D. G. (1968). Natural language input for a computer problem -solving system. In M. Minsky (Ed.), Semantic Information Processing (Chapter 3). Cambridge, MA: MIT Press.

Carpenter, P., & Just, M. (1987). The role of working memory in comprehension (To be published in 1988). Proceedings of the

Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psychology Department, Pittsburgh, PA.

Champernowne, D. G. (1953). A model of income distribution. Economic Journal, 63, 318-351.

Charness, N. (1987). Expertise in chess and bridge (To be published in 1988). Proceedings of the Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psychology Department, Pittsburgh, PA.

Crecine, J. P. (1969). Governmental problem-solving: A computer simulation of municipal budgeting. Chicago, IL: Rand-McNally.

Ericsson, A., & Staszewski, J. (1987). Skilled memory and expertise: Mechanisms of exceptional performance (To be published in 1988). Proceedings of the Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psycholgy Department, Pittsburgh, PA.

Goodwin, R. M. (1947). Dynamical coupling with especial reference to markets having production lags. Econometrica, 15, 181-204.

Hadamard, J. (1945). The psychology of invention in the mathematical field. Princeton: Princeton University Press.

Hayes, J. R. (1987). Memory organization and world-class performance (To be published in 1988). Proceedings of the Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psychology Department, Pittsburgh, PA.

Ijiri, Y. & Simon, H. A. (1977). Skew distributions and the sizes of business firms. Amsterdam: North Holland Publishing Company.

Johnson-Laird, P. N. (1983). Mental models. Cambridge, MA: Harvard University Press.

Klahr, D., & Dunbar, K. (1987). Developmental differences in scientific discovery strategies (To be published in 1988). Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psychology Department, Pittsburgh, PA.

Korolyuk, V. S., Polischuk, L. I., & Tomusyak, A. S. (1969). A limit theorem for semi-markov processes (In Russian). Kibernetika, 5, 144-145.

Kosslyn, S. M. (1980). Image and mind. Cambridge, MA: Harvard University Press.

Kotovsky, K., & Fallside, D. (1987). Representation and transfer in problem solving (To be published in 1988). Proceedings of the Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-

Mellon University, Psychology Department, Pittsburgh, PA.

Kulkarni, D. & Simon, H. A. The processes of scientific discovery: The strategy of experimentation. Unpublished manuscript, Carnegie-Mellon University, Department of Computer Science, Research papers 86-111.

Langley, P. W., Simon, H. A., Bradshaw, G., & Zytkow, J. (1987). Scientific discovery: Computational explorations of the creative processes. Cambridge, MA: MIT Press.

Larkin, J. H. (1987). Display-based problem solving (To be published in 1988). Proceedings of the Twenty-First Carnegie-Mellon Symposium on Cognition. Carnegie-Mellon University, Psychology Department, Pittsburgh, PA.

Larkin, J. H., McDermott, J., Simon, D., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. Science, 208, 1335-1342.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth 10,000 words. Cognitive Science, 11, 65-100.

Lotka, A. J. (1924). Elements of physical biology. Baltimore, MD: Williams and Wilkins.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. In Willis Jackson (Ed.), Communication Theory. (pp. 486-502). London: Butterworths.

McCorduck, P. (1979). Machines who think. San Francisco, CA: W. H. Freeman.

Miller, G. A., & Johnson-Laird, P. N. (1976). Language and perception. Cambridge, MA: Harvard University Press.

Nevell, A. (1973). You can't play 20 questions with Nature and win. In William G. Chase (Ed.), Visual Information Processing, New York: Academic Press.

Nevell, A., & Simon, H. A. (1956). The logic theory machine. IRE Transactions on Information Theory, IT-2, 3, 61-79.

Nevell, A., & Simon, H. A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Novak, G. S. Jr. (1976). Computer understanding of physics problems stated in natural language. Technical Report NL-30, Department of Computer Sciences, University of Texas, Austin, Texas.

Peirce, B. O. (1929). A short table of integrals (third rev. ed.). Boston, MA: Ginn & Company.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain. Psychological Bulletin, 80, 1-24.

Simon, H. A. (1935). Administration of public recreational facilities in Milwaukee. Unpublished manuscript, (Quoted in Administrative Behavior. (pp. 211-212).

Simon, H. A. (1947). Administrative behavior. New York: Macmillan.

Simon, H. A. (1955). A behavioral model of rational choice. Quarterly Journal of Economics, 69, 99-118.

Simon, H. A. (1955). On a class of skew distribution functions. Biometrika, 52, 425-440.

Simon, H. A., & Ando, A. (1961). Aggregation of variables in dynamic systems. Econometrica, 29, 111-138.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, H. A. (1979). Models of thought. New Haven, CT: Yale University Press.

Simon, H. A. (1981). The Sciences of the artificial. Cambridge, MA: MIT Press.

Simon, H. A. (1982). Models of bounded rationality (Volumes 1-2). Cambridge, MA: MIT Press.

Yu, B., Zhang, W., Jing, Q., Peng, R., Simon, H., & Zhang, G. (1985). STM capacity for Chinese and English language materials. Memory and Cognition, 13, 202-207.

Yule, G. U. (1924). A mathematical theory of evolution, Based on the conclusions of Dr. J. C. Willis, F.R.S. Philosophical Transactions, 21-83.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. Memory and Cognition, 13, 193-201.